IET Image Processing

Research Article

Dense optical flow based background subtraction technique for object segmentation in moving camera environment

ISSN 1751-9659
Received on 2nd August 2019
Revised 17th May 2020
Accepted on 13th July 2020
doi: 10.1049/iet-ipr.2019.0960
www.ietdl.org

Arati Kushwaha[1], Ashish Khare[1] ✉, Om Prakash[2], Manish Khare[3]
[1]Department of Electronics and Communication, University of Allahabad, Allahabad, India
[2]Department of Computer Science and Engineering, H.N.B. Garhwal University, Srinagar (Garhwal), India
[3]Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India
✉ *E-mail: khare@allduniv.ac.in*

Abstract: Segmentation of moving object in video with moving background is a challenging problem and it becomes more difficult with varying illumination. The authors propose a dense optical flow-based background subtraction technique for object segmentation. The proposed technique is fast and reliable for segmentation of moving objects in realistic unconstrained videos. In the proposed work, they stabilise the camera motion by computing homography matrix, then they perform statistical background modelling using single Gaussian background modelling approach. Moving pixels are identified using dense optical flow in the background modelled scenario. The dense optical flow provides motion information of each pixel between consecutive frames, therefore for moving pixel identification they compute motion flow vector of each pixel between consecutive frames. To distinguish between foreground and background pixels, they labelled each pixel and thresholding the magnitude of motion flow vector identifies the moving pixels. The effectiveness of the proposed algorithm has been evaluated both qualitatively and quantitatively. The proposed algorithm has been evaluated on several realistic videos of different complex conditions. To assess the performance of the proposed work, the authors compared their algorithm with other state-of-art methods and found that the proposed method outperforms the other methods.

1 Introduction

In past two decades, computer vision community has been working on different areas like robotics, automated surveillance, vehicular traffic analysis, human–computer interaction, human activity recognition and so on, mainly to deal with safety and security [1] concerns. Moving object segmentation is one of the important steps in computer vision applications. Segmentation of moving object by moving camera such as pan-tilt-zoom (PTZ) camera or hand-held camera is a challenging and difficult task due to varying background along with the motion of camera. Moving object segmentation is a process of detection of intended object and/or detection of foreground pixels in video frames. The segmentation of moving object is a difficult task [2, 3] due to the following reasons:

(i) 2D scene is projection of 3D world.
(ii) Presence of noise and blur.
(iii) Complex motion of object.
(iv) Camera motion.
(v) Object occlusion.
(vi) Varying object shape from frame to frame.
(vii) Dynamic background.
(viii) Varying illumination condition.

In the present work, we propose a dense flow based moving object segmentation method in moving camera environment. The segmentation of moving object was commonly performed by the background subtraction techniques [4–11]. The background subtraction techniques perform subtraction between background and current frame. These methods performed poor segmentation in case of slow motion of object, abrupt change in illumination, complex background which were not well suited for realistic unconstrained videos. To address these problems, we propose a dense optical flow based background subtraction approach in moving camera environment. We integrated single Gaussian based background subtraction technique with dense optical flow method to compute motion flow vector. The main motivation behind use of dense optical flow [12–15] for computation of motion flow vector is that background pixels have negligible motion as compared to foreground pixels [16]. Firstly, camera motion was estimated and then background motion has been stabilised to get frame sequence. These frame sequences could be treated as frames in videos recorded with a static camera. For videos with varying illumination and slow object motion, false labelling problem is the major issue in object segmentation [16]. To overcome the false labelling problem, we use motion vector of each pixel. Since, in stabilised camera motion and background modelling, each background pixel have negligible motion than the moving pixels. The proposed object segmentation approach consists of following subtasks:

(i) Stabilisation of camera motion by computing overlapped region between current frame and background.
(ii) Background modelling using single Gaussian modelling approach with varying learning rate.
(iii) Computation of motion flow vector of each pixel, using dense optical flow between current frame and background.
(iv) Segmentation of moving object by subtracting the background frame from current frame followed by thresholding the magnitude of motion flow vectors of each frame.

The performance of proposed method is evaluated and compared with other state-of-the-art methods. The proposed method is compared with the nine well-established methods for moving object segmentation in moving camera environment. The methods used in comparison are as follows:

(i) Moving object segmentation based on scene conditional background updation approach, proposed by Yun *et al.* [17].
(ii) Moving object segmentation based on graph cut approach, proposed by Zhou *et al.* [10].
(iii) Moving object segmentation based on stochastic approximation, proposed by Lopez *et al.* [6].
(iv) Moving object segmentation based on block-based background modelling approach, proposed by Yi *et al.* [8].

(v) Moving object segmentation based on spatio-temporal background modelling approach, proposed by Kim *et al.* [4].
(vi) Moving object detection in a video grabbed by unstable camera, proposed by Lee *et al.* [11].
(vii) A robust single and multiple moving object detection, tracking and classification methods proposed by Mahalingam *et al.* [18].
(viii) Moving object detection for a moving camera based on global motion compensation and adaptive background model, proposed by Yu *et al.* [19].
(ix) Moving object detection based on optical flow estimation and Gaussian mixture model, proposed by Cho *et al.* [20].

The performance of the proposed method was not only assessed visually but also quantitatively. For quantitative performance evaluation, we used six different evaluation measures viz. percentage of correct classification (PCC), Jaccard coefficients (JC), figure of merit (FM), quality percentage (QP), average difference (AD) and mean square error (MSE).

The rest of the paper is organised as follows. Literature review is given in Section 2. An overview of different algorithms used in different stages of proposed work has been discussed in Section 3. Section 4 presents the proposed segmentation approach. Experimental results and conclusions are given in Section 5 and Section 6, respectively.

2 Literature review

Several methods have been proposed for moving object segmentation in literature. In the literature, the segmentation of moving object has been mainly performed using three different classes of approaches. These approaches can be categorised as

(i) Object modelling based approach [17, 21, 22].
(ii) Optical flow based approach [16, 23, 24].
(iii) Background subtraction based approach [25–28].

In object modelling based approach [17, 21, 22], target objects are extracted by applying object segmentation techniques. The methods based on this approach mainly focuses on appearance and motion coherence of moving pixels. These methods perform well when object is large and visually noticeable. This class of methods are not suitable for segmentation of tiny objects. Also, these methods are computationally costly. However, the real videos may contain objects of very small to that of large size.

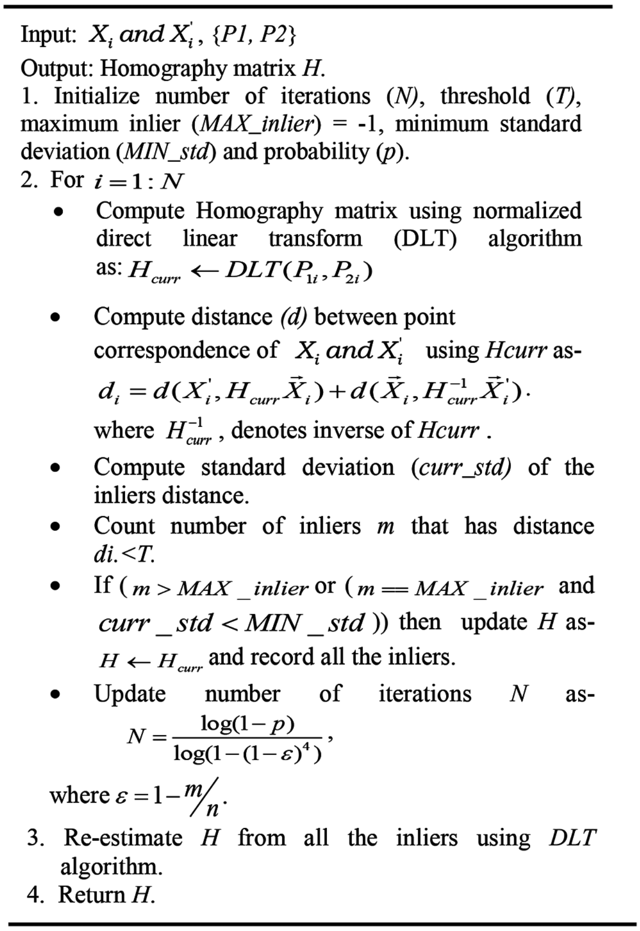
In optical flow-based approaches [16, 23, 24], firstly, the consecutive frames are aligned by using transformation matrices such as affine transform, homography matrix and so on, then motion flow vectors between aligned and current frame are computed. Moving foreground pixels are identified by thresholding the magnitude of motion flow vectors. The methods based on optical flow perform well in case of simple background and single object but not suitable for complex background and multiple objects.

In background subtraction-based approaches [18, 25–29], moving objects are extracted by subtraction of background from the current frame. In these methods, it is assumed that pixel intensity is same over long period of time, as in static camera scenario. From past few decades, researches are working on background subtraction techniques due to their low complexity and achieved good results. However, these techniques are not suitable for varying background videos such as videos taken by PTZ and unmanned aerial vehicles cameras and so on.

Therefore, considering the above facts, efficient background modelling is required for background subtraction. For background modelling, two types of approaches based on panoramic background modelling and non-panoramic background modelling have been used. In panoramic background modelling based approach [30–33], a large panorama is constructed from entire view of camera motion, by stitching input frames using image registration [34, 35], such as tonal mosaic alignment [34], Lucas–Kanade tracking [35] (LKT), scale invariant feature transform (SIFT) [36] and so on. After generating the large panorama background, moving object can be detected by applying subtraction between current frame and corresponding region in panorama background. Mittal and Huttenlocher [30] proposed background modelling method for moving camera environment and moving objects are segmented by matching Gaussian model. The idea of construction of large panorama followed by frame difference for object segmentation is exploited by Cucchiara *et al.* [31]. Methods proposed by Cho and Kang [32] and Xue *et al.* [33] have constructed panorama, using all possible view of camera, using different motion model and moving objects are segmented by existing segmentation approaches. Panoramic background methods have an advantage that can use any segmentation technique after panoramic background modelling. These methods suffer from shortcomings like registration error, parallax effect, slow initialisation and large computation time and memory requirement [4].

Shortcomings of panoramic background modelling techniques can be avoided by using non-panoramic background modelling techniques [4–11, 16, 17, 19, 20]. In this approach, background has been modelled in the size of input frames, to avoid construction of large panorama. After that, foreground has been detected by subtracting background from current frame. Normally background is in motion due to camera motion. In these approaches camera motion has been stabilised by computing a global transform matrix. The camera motion has been compensated by warping background model with current input frame. Kim *et al.* [4] proposed a spatio-temporal background modelling-based approach with varying learning rate for moving object segmentation. Hu *et al.* [5] used an integrative approach for moving object detection based on foreground feature point and foreground regions. Lopez *et al.* [6] has considered moving camera segmentation approach using stochastic approximation learning. Kim *et al.* [7] proposed a method based on clustering of optical flow vector for foreground and background pixel identification. Yi *et al.* [8] exploited block-based background modelling for moving object segmentation in moving background environment. Zhou *et al.* [10] proposed an approach for moving object segmentation in urban environment, based on graph-cut framework. Lee *et al.* [11] used kernel density estimation and difference of Gaussian approach for moving object detection after global camera motion stabilisation. An adaptive background modelling approach has been proposed by Yu *et al.* [19] in which local pixel difference and local changes between current frame and background frame has been used. Cho *et al.* [20] has proposed a method for moving object segmentation, in moving camera environment, which uses optical flow estimation for camera motion stabilisation and then Gaussian mixture model has been used for object segmentation. The above-mentioned methods resolve shortcomings of panoramic background approaches but non-panoramic background modelling based approaches still suffers from wrong detection of background pixels as foreground pixels (false labelling). This wrong detection is due to slow motion of object than background, abrupt change in illumination [16, 17]. Kurnianggoro *et al.* [16] used dense optical flow based approach for moving object segmentation, in moving camera environment based on the assumption that foreground pixels have larger magnitude and background pixels have negligible magnitude. Upon camera motion stabilisation, camera motion pixels could be brought in rest [4, 25, 26]. An adaptive background model updation is required for effective and accurate segmentation, but these approaches had not considered any background updation scheme and this is the reason why it results in poor segmentation in presence of view point changes, complex background and varying object size. Yun *et al.* [17] proposed a background modelling technique for moving object detection in moving camera environment based on scene changes. In this work, the problem of false labelling of background pixels as foreground pixel has been solved up to certain extent by background updation based on the scene changes, but this method is still not giving good segmentation results in presence of complex background in realistic unconstrained videos.

Thus, background subtraction based methods in moving camera environment are suffering from various problems like false labelling of background pixel as moving object pixels, abrupt change in illumination and compensation error and so on. Other



```
Input: $X_i$ and $X_i^{'}$, {P1, P2}
Output: Homography matrix H.
1. Initialize number of iterations (N), threshold (T), maximum inlier (MAX_inlier) = -1, minimum standard deviation (MIN_std) and probability (p).
2. For $i = 1 : N$
    • Compute Homography matrix using normalized direct linear transform (DLT) algorithm as: $H_{curr} \leftarrow DLT(P_{1i}, P_{2i})$
    • Compute distance (d) between point correspondence of $X_i$ and $X_i^{'}$ using Hcurr as-
      $d_i = d(X_i^{'}, H_{curr}\vec{X}_i) + d(\vec{X}_i, H_{curr}^{-1}\vec{X}_i^{'})$.
      where $H_{curr}^{-1}$, denotes inverse of Hcurr .
    • Compute standard deviation (curr_std) of the inliers distance.
    • Count number of inliers m that has distance di.<T.
    • If ( $m > MAX\_inlier$ or ( $m == MAX\_inlier$ and $curr\_std < MIN\_std$ )) then update H as- $H \leftarrow H_{curr}$ and record all the inliers.
    • Update number of iterations N as- $N = \frac{\log(1-p)}{\log(1-(1-\varepsilon)^4)}$,
  where $\varepsilon = 1 - m/n$.
3. Re-estimate H from all the inliers using DLT algorithm.
4. Return H.
```

Fig. 1 *Algorithm for computation of homography matrix*

limitations of these methods are object occlusion, background clutter, loss of information due to 3D world projection into 2D scenes, complex and dynamic background, nearly similar intensity values of foreground and background, complex motion of object and so on.

3 Object segmentation in moving camera environment

A conventional background subtraction technique fails to extract moving object in presence of camera motion as well as object motion. The work presented in this paper is based on global camera motion stabilisation for object segmentation in moving camera environment. In this section, we present an overview of different algorithms used in different stages of the proposed work in separate subsections.

3.1 Homography matrix

Homography matrix is a 3×3 projective transform matrix of a plane, which gives point-to-point correspondence between two frames captured by moving camera [37]. To compute homography matrix between consecutive frames, at least four corresponding points are needed between them. We have used RANSAC (RANdom SAmple Consensus) algorithm [38] for computation of homography matrix. RANSAC is an iterative model for selecting inliers, among many outliers, that are compatible with homography between consecutive frames. Homography matrix is computed globally to stabilise camera motion by using extracted feature points from background frame and their corresponding points from current frame. It can be given as follows:

$$\boldsymbol{H} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \qquad (1)$$

where h_{11}, h_{12}, h_{13},…h_{33} are nine projective points between two images of 3×3 projective transform matrix.

The algorithm for computation of homography matrix using RANSAC is given in Fig. 1. In the algorithm, $\{P_1, P_2\}$ are set of initial corresponding feature points (computed using Harris corner matching) between two consecutive frames X_i and X_i'.

3.2 Camera motion compensation

Extraction of moving object in moving camera environment is not an easy task due to the presence of background. Therefore, first of all we need to stabilise the camera motion and then by applying background subtraction technique, moving object can be extracted well. LKT is used to compute displacement of keypoints between consecutive frames. Camera motion is estimated by finding mathematical relation that maps pixel co-ordinates from one frame to the other as follows:

$$F_i = \boldsymbol{H} \cdot F_i' \qquad (2)$$

where F_i is ith feature extracted from the background frame, F' is ith tracked feature in current frame at the corresponding location and $\boldsymbol{H}$ is homography matrix. By multiplying homography matrix with background model, overlapped and newly covered region in the current frame are obtained, and the motion of camera is stabilised. Newly covered region are directly labelled as background area and are updated in next iteration.

3.3 Background modelling

To model the background, as in fixed camera case, initial frame is considered as background frame. In this work, spatio-temporal single Gaussian background modelling technique is used to model the background [4, 25–27], i.e. single variance and single mean have been taken for each pixel. Background is updated with those pixels, which are previously labelled as background, with varying learning rate. Initially, the learning rate is taken as 1 for each background pixel and as camera moves to next frame, its value increases for overlapped region and for every newly covered region learning rate again taken as 1. Background pixel labels are decided by computing motion vectors of each pixels, as well as displacement between current pixel with neighbourhood pixel of background frame. Let x_c is pixel of current frame and x_b is pixel of background frame at point (x,y), then backgrounds are updated using updated mean and variances

$$\mu_t(x_b) = (1 - \alpha) \cdot \mu_{t-1}(x_b) + \alpha \cdot I(x_c) \qquad (3)$$

$$\sigma_t^2(x_b) = (1 - \alpha) \cdot \sigma_{t-1}^2(x_b) + \alpha \cdot (I(x_c) - \mu_t(x_b))^2 \qquad (4)$$

$$\alpha = \frac{1}{L(x_b)} \qquad (5)$$

where $\mu_t(.)$ and $\sigma_{2t}(.)$ are mean and variance of background at time t. $L(\cdot)$ is learning rate.

As discussed in background modelling and updation technique, learning rate is varying and it increases with motion of camera for overlapped region. The overlapped regions are directly labelled as background pixel [4]. This labelling yields extraction of motion information of each pixel which is helpful to avoid false labelling of motionless background pixels.

3.4 Foreground and background pixel identification

After camera motion stabilisation, foreground pixels are identified by subtracting the background frame from the overlapped region of current frame. The conventional approach of segmentation of moving object suffers from problems like registration error, parallax effect, slow initialisation, huge memory uses, abrupt change in illumination and false labelling of moving pixels by background pixel. Here in the proposed work, the shortcomings like registration error, slow initialisation, huge memory uses and

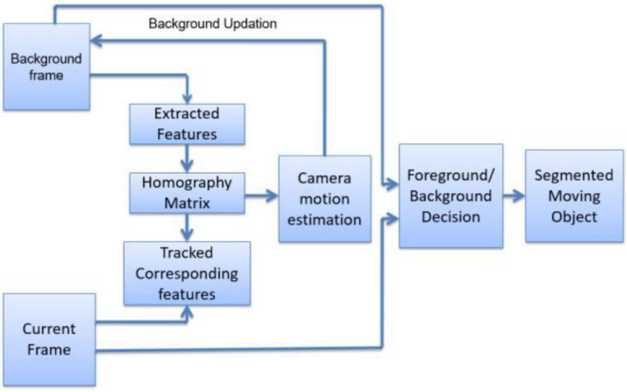


Fig. 2 *Block diagram of the proposed approach*

parallax effect have been resolved by modelling background in the size of current frame. Since background and foreground both have different velocities, therefore the problem of abrupt change in illumination and false detection of moving pixels as background pixels have been resolved by computing motion flow vector of each pixel. Moving pixels are identified by computing labels for deciding background and foreground pixels by comparing current frame with background model followed by thresholding the motion vector of each pixel as follows.

The label for the decision of foreground\background is calculated as

$$D = \frac{1}{\sigma_{t-1}^2(x_b)} \cdot (I(x_c) - \mu_t(x_b))^2 \qquad (6)$$

Magnitude of motion vectors for each pixel is computed as

$$M = \sqrt{\boldsymbol{M}_x^2 + \boldsymbol{M}_y^2} \qquad (7)$$

where $\boldsymbol{M}_x$ and $\boldsymbol{M}_y$ are optical flow vectors in x and y directions.

Foreground and background pixels are labelled as

$$\text{If } (D < \beta) \qquad (8)$$

$$L(x_c) = \begin{cases} \text{foreground} & \text{if}(M > \tau) \\ \text{background} & \text{otherwise} \end{cases} \qquad (9)$$

where $\mu_t(\cdot)$ and $\sigma_t^2(\cdot)$ are mean and variance of background model at time t, β and τ are thresholds used for identification of background and foreground pixels, respectively, and $I(x_c)$ is intensity value of the current frame. Foreground and background pixels are identified based on labels of pixels followed by thresholding of motion flow vectors which are computed using dense optical flow approach. Motion flow vectors help in avoiding the false identification of motionless pixel of background. This is due to negligible magnitude of motion vector of background pixels compared to moving pixels of foreground. However, the segmentation results may be affected by the presence of noise and nearly same intensity levels of neighbouring pixels of object. To overcome this problem results could be further refined by post-processing morphological operation. In this work, we used dilation and erosion morphological operators to refine the result with diagonal and ellipse structuring elements of size 3 × 3 and 5 × 5.

3.5 Dense optical flow

Optical flow computes the vector fields that determine velocity of each moving pixel in time-varying image sequences [12, 15]. Dense optical flow captures the motion of each pixel in consecutive frames, i.e. it shows displacement of intensity pattern between consecutive frames in the form of point-to-point correspondence [12–15, 39, 40] of pixels. There are several approaches for motion estimation through optical flow such as correlation or block matching [15], feature tracking [39], energy-based approaches [39], pixel intensity-based approaches (brightness remains consistent with time) [13], gradient-based approaches [12] and polynomial expansion transform-based approaches [40]. In this work, we use polynomial expansion transform-based approach proposed by Gunnar Farneback [40]. The first step of this approach is to approximate the neighbourhood of the consecutive frames using polynomial expansion transform. Then displacement vectors are estimated using these polynomial expansions after number of refinements [40].

4 Proposed method

Objective of this work is to present a fast and reliable method for object segmentation in moving camera environment for realistic and unconstrained videos. Object segmentation in moving camera environment is not easy tasks due to the presence of two types of motion – background motion and object motion. Therefore, first task is to stabilise camera motion, so that the frame sequences could be treated as in the case of static camera. After stabilisation of camera motion, single Gaussian background modelling approach is used for background modelling that results in improved segmentation at low computational cost.

Hybridisation of existing background subtraction technique with camera motion compensation still suffers from slow initialisation, abrupt change in illumination, appearance changes, compensation error and false labelling problem [16, 17], as discussed in Sections 2 and 3. To handle these problems, a new approach using dense optical flow has been proposed for differentiating between foreground and background pixels. Computation of labels corresponding to each pixel as foreground or background pixel, if we revalidate each pixel as a foreground or background pixel through motion vector of each pixel, we can resolve the problem of illumination change and false labelling. This is due to a larger magnitude of the object pixels compared to the background pixels in motion flow vector. Therefore, use of dense optical flow, the motion flow vector of each pixel has been computed for the correct identification of moving object pixels. The block diagram of the proposed method is shown in Fig. 2. Step-wise illustration of Fig. 2 is as follows:

(i) Initialise the first frame as background frame and update it in each iteration with new frame after camera motion estimation, using single Gaussian background modelling technique.

(ii) Extract the feature vector from the background frame. Harris corner feature [41] has been used as feature due to its invariance nature w.r.t. rotation and translation, and varying illumination condition [42].

(iii) Track the corresponding extracted feature in the current frame using LKT [34] to register two consecutive frames.

(iv) Compute homography matrix [37] for computing point correspondences in consecutive frames.

(v) Stabilise motion of camera by computing overlapped region and newly covered region by warping background with current frame.

(vi) To identify object pixels among foreground and background pixels, compute motion flow vector of each pixel between background and current frame. Foreground and background pixels are differentiated by subtracting background from current frame on the basis of computed labels [26] followed by thresholding the magnitude of motion flow vector of each pixel.

(vii) Results obtained are further processed by morphological operation to refine the extracted moving objects (see Fig. 3).

5 Experimental results and discussion

The proposed method of moving object segmentation in moving camera scenario is implemented in C++ environment with computer vision open source library 3.0. We performed experiments on number of realistic unconstrained videos recorded in various real scenarios like cloudy, fog and rainy weather. We tested our method on five self-created datasets (https://drive.google.com/open?

```
Input:  Video clip of size m×n×t
Output: Foreground mask fg
  (i).    Let Vi be $i^{th}$ input video clip consisting of n number of frames, $I_c$ is current frame and $I_b$ is
          first frame initialized as background frame in the $i^{th}$ video clip.
  (ii).   For i=1 to n
  (iii).  $P_1(p_1, p_2, p_3......p_n) \leftarrow Harris\,Corner(I_b)$ •
          Track the feature points corresponding to P1 as follows: $P_2(p'_1, p'_2, p'_3......p'_n) \leftarrow LKT(I_c)$
  (iv).   Compute homography matrix using points (P1, P2) using RANSAC algorithm as discussed in section 3.1.
  (v).    Perform warping of background with current frame to estimate camera motion as
              $I_b = H \cdot I_c$ ,
          where H is Harris corner features.
  (vi).   Let, $\mu_t(\cdot)$ and $\sigma^2_t(\cdot)$ are mean and variance of background frame at time t and $L(\cdot)$ is learning
          rate. Single Gaussian background modelling approach has been used for background
          modelling. $\mu_t(\cdot)$ and $\sigma^2_t(\cdot)$ are updated as-
              $\mu_t(x_b) = (1-\alpha)\cdot\mu_{t-1}(x_b) + \alpha\cdot I(x_c)$
              $\sigma^2_t(x_b) = (1-\alpha)\cdot\sigma^2_{t-1}(x_b) + \alpha\cdot(I(x_c) - \mu_t(x_b))^2$
          where, $\alpha = \frac{1}{L(x_b)}$
  (vii).  Compute label for foreground/background pixel as
              $D = \frac{1}{\sigma^2_{t-1}(x_b)}\cdot(I(x_c) - \mu_t(x_b))^2$
          where, $x_c$ is pixel of current frame, $x_b$ is pixel of background frame at point (x,y)
  (viii). Let, Mx and My are motion vectors in x and y directions. Magnitude of motion flow vectors are
          computed as
              $M \leftarrow \sqrt{M_x{}^2 + M_y{}^2}$
  (ix).   Foreground/Background decision is made as-
          if (D < β)
              $L(x_c) = \begin{cases} foreground & if\,(M > \tau) \\ background & otherwise \end{cases}$
          end if
  (x).    Process resulted foreground by morphological operator to refine the result.
  (xi).   end for
```

Fig. 3 *Algorithm for proposed moving object segmentation*

id=1o_37sfIDOaO3HW99TwMYVWLT_tT3CeJ7) and one standard dataset used by Yun *et al.* [17] (http://pil.snu.ac.kr/user/kmyun/file/scbudata.zip). The dataset [17] consists of various movie clips of dancing, skating, swing, walking and so on. For representative purpose, in the paper, results on three datasets are presented. However, all the results have been uploaded on the website (https://drive.google.com/file/d/1LhTz6pWanA67byapu2b5EJWbpYB1FJEu). The proposed method is compared with other state-of-art methods proposed by Yun *et al.* [17], Zhou *et al.* [10], Lopez *et al.* [6], Yi *et al.* [8], Kim *et al.* [4], Lee *et al.* [11], Mahalingam *et al.* [18], Yu *et al.* [19] and Cho *et al.* [20] for all representative video sequences.

The segmentation results on some representative frames by the proposed method and other methods used in comparison have been shown in Figs. 4–6. A complete set of results on different videos are available at https://drive.google.com/file/d/1LhTz6pWanA67byapu2b5EJWbpYB1FJEu. The qualitative results are not sufficient to judge the performance of the segmentation. Therefore, we have also evaluated the performance quantitatively. For quantitative evaluation, we have constructed ground truth by publicly available tool Virtual Dub (http://www.virtualdub.org/index.html). In this paper, we used six standard evaluation metrics – PCC, JC, FM, QP, AD and MSE [9, 29, 43]. The metric values on different datasets have been given in Tables 1–3.

Let TP is number of true positive pixels, FP is number of false positive pixels, TN is number of true negative pixels and FN is number of false negative pixels. The computational equations of different quantitative measures used in the proposed work are given below.

PCC [43], which is also stated as accuracy, gives the total number of correct predictions. Higher value of PCC shows good segmentation result. PCC can be computed [43] as

$$\mathrm{PCC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP}} \tag{10}$$

JC [9] is defined as

$$\mathrm{JC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{11}$$

Higher value of JC shows better segmentation results.

FM [43] is defined as

$$F = \frac{2 \times R \times P}{R + P} \tag{12}$$

where *R* and *P* are recall and precision, respectively

$$\mathrm{Recall}(R) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{13}$$

$$\mathrm{Precision}(P) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{14}$$

Higher value of FM shows good segmentation results.

QP [29] is computed as

$$\mathrm{QP} = 100 \times \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{15}$$

The higher value of QP shows good segmentation results.

AD [29] between ground truth and segmented frame is defined as

$$\mathrm{AD} = \frac{\sum_{i=1}^{x}\sum_{j=1}^{y}(G_{i,j} - S'_{i,j})}{xy} \tag{16}$$

where $G_{i,j}$ is ground truth frame and $S'_{i,j}$ is segmented frame of dimension $x \times y$. The smaller value of AD shows better segmentation results.

MSE [9] between ground truth frame and segmented frame can be computed as

$$\mathrm{MSE} = \frac{\sum_{i=1}^{x}\sum_{j=1}^{y}(G_{i,j} - S'_{i,j})^2}{xy} \tag{17}$$

where $G_{i,j}$ is ground truth frame and $S'_{i,j}$ is segmented frame of dimension $x \times y$. Lower the value of MSE, better is the segmentation.

5.1 Experiment #1

In this experiment, we used a video clip containing 1140 frames of size 640 × 360 and frame rate 24 fps. The input frames and segmented results of the proposed method and other methods used in comparison have been presented. The methods used in comparison are Yun *et al.* [17], Zhou *et al.* [10], Lopez *et al.* [6], Yi *et al.* [8], Kim *et al.* [4], Lee *et al.* [11], Mahalingam *et al.* [18], Yu *et al.* [19] and Cho *et al.* [20]. The results for few representative frame nos. 50, 450, 725 and 925 have been given in Fig. 4. From Fig. 4, we can observe that the proposed method performed better for the objects of size ranging from very small (Fig. 4*c*) to that of large size (Figs. 4*a* and *b*). Also, it is clear that the proposed method is able to segment the objects in the scenes with fog and rainy weather environment. Although, we have not received results as accurate as ground truth. The results have also been evaluated quantitatively using the performance metrics – PCC, JC, FM, QP, AD and MSE [9, 29, 43]. These metric values have been shown in Table 1. The best values are shown in bold. From Fig. 4, it can be observed on frame no. 50 that the methods proposed by Zhou *et al.* [10], Lee *et al.* [11], Yu *et al.* [19] and Cho *et al.* [20] were failed to segment any portion of the object [Fig. 4*a*, from left first row fifth column, Fig. 4*a*, from left second row third column, Fig. 4*a*, from left second row fifth column and Fig. 4*a*, from left second row sixth column]. However, the methods Yun *et al.* [17], Lopez *et al.* [6], Yi *et al.* [8], Kim *et al.* [4] and Mahalingam *et al.* [18] performed poorly in segmenting the object in frame no. 50

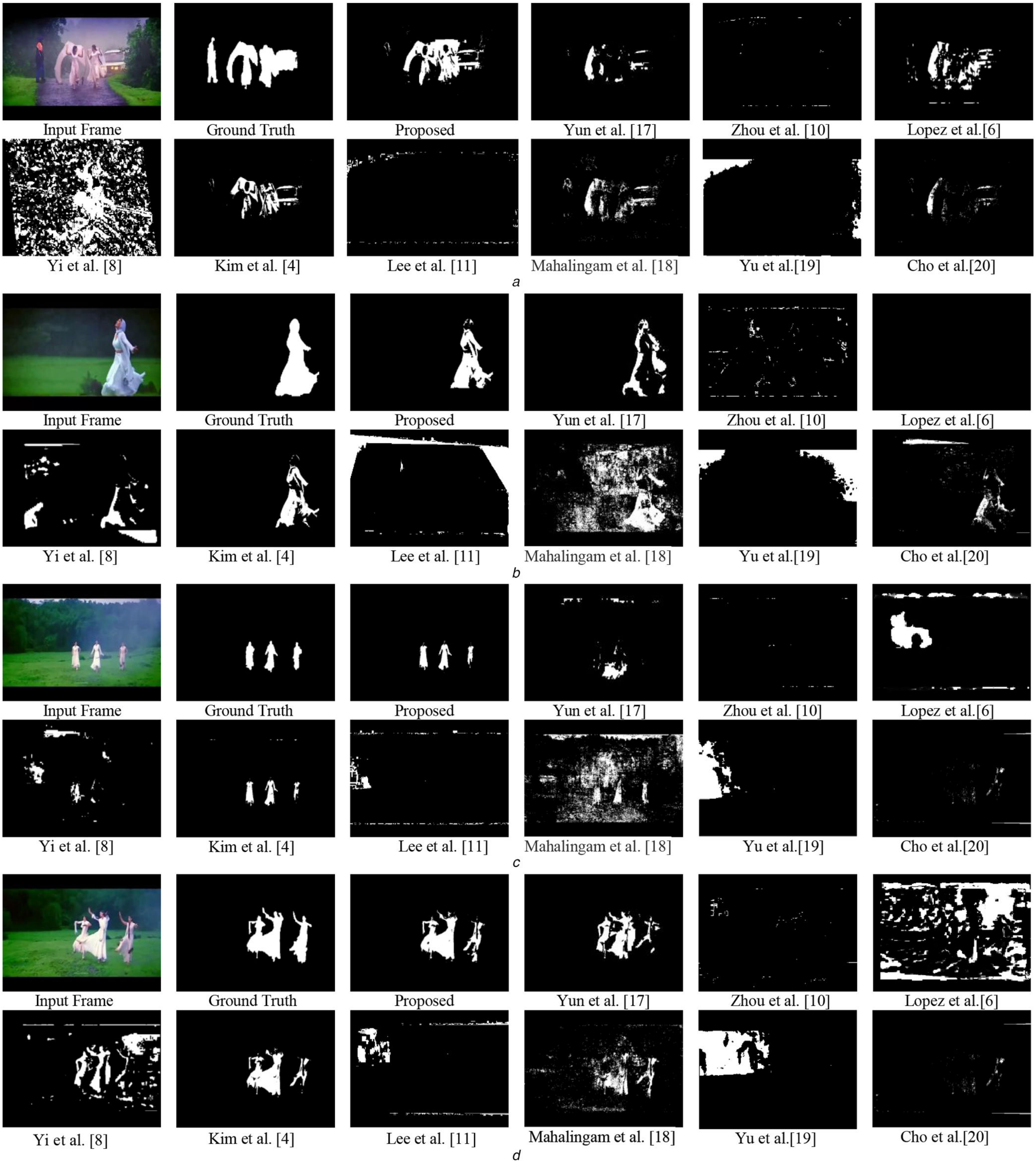


Fig. 4 *Segmentation results on a video clip of frame size 640 × 360, frames are arranged from top to bottom as*
(a) Frame no. 50, ***(b)*** Frame no. 450, ***(c)*** Frame no.725, ***(d)*** Frame no. 925

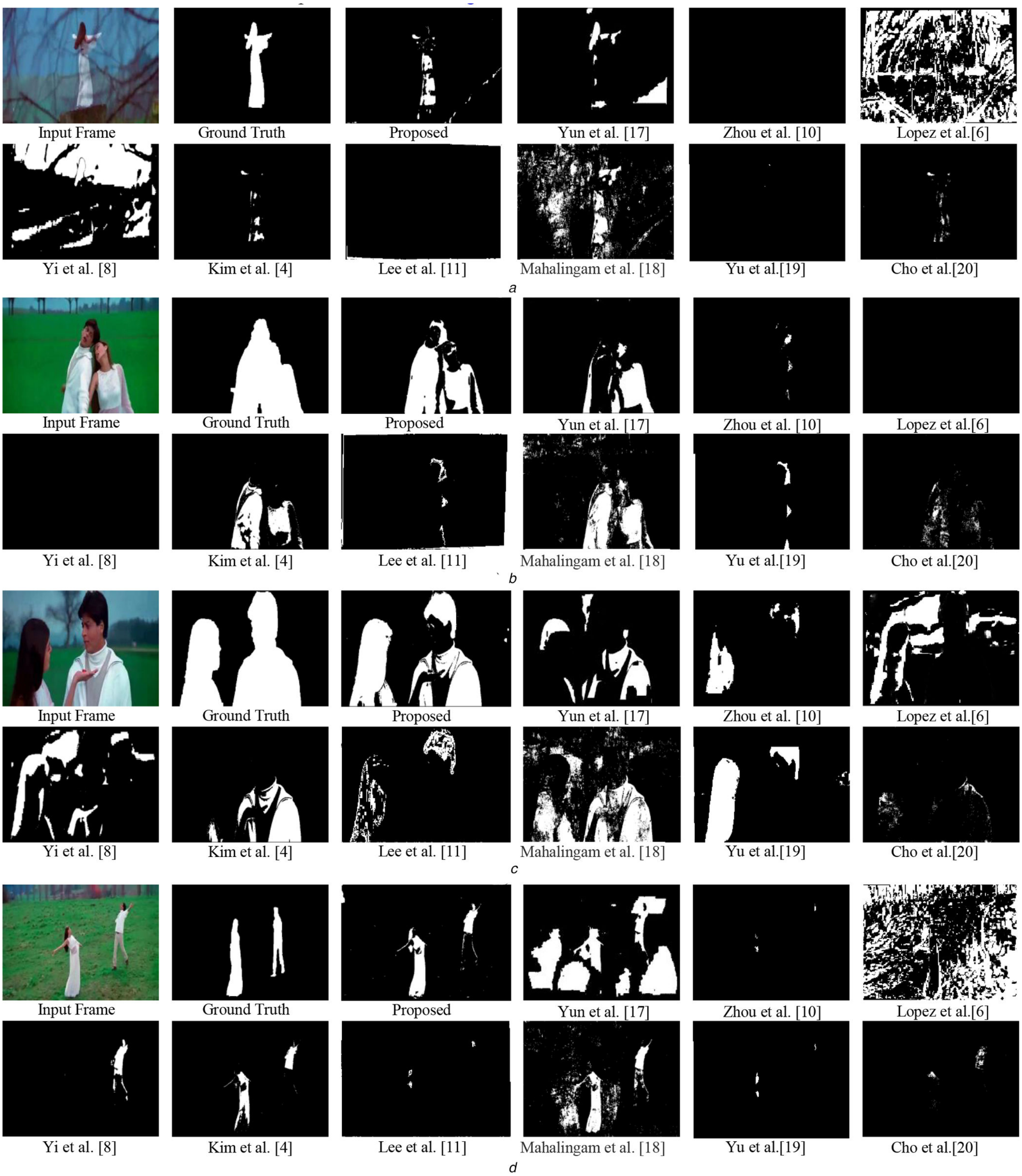


Fig. 5 *Segmentation results on a video clip of frame size 640 × 272, frames are arranged from top to bottom as*
(a) Frame no. 120, ***(b)*** Frame no. 450, ***(c)*** Frame no. 950, ***(d)*** Frame no. 1060

(Fig. 4*a*). The segmentation results on frame no. 450 are shown in Fig. 4*b*. We can see from Fig. 4*b* that the methods Zhou *et al.* [10], Lopez *et al.* [6], Lee *et al.* [11] and Yu *et al.* [19] could not perform any segmentation while the methods Yun *et al.* [17], Lopez *et al.* [6], Yi *et al.* [8], Kim *et al.* [4], Lee *et al.* [11] and Cho *et al.* [20] produced poor segmentation. In representative frame nos. 725 and 925, we observed almost similar segmentation performance of the methods as on frame no. 450. Thus, overall, from Fig. 4, we can see that the proposed method produces better segmentation results in every frame.

To assess the performance of the segmentation methods, the visual results are not sufficient and hence we have also used quantitative performance analysis for judging the performance of the segmentation capability of the proposed method. We computed six performance measures for the proposed method along with methods used in comparison. These metric values are given in Table 1. From Table 1, we can observe that the proposed method has higher metric values (PCC, JC, FM, QP), lower MSE and AD values in all cases except the method proposed by Yu *et al.* [19] that results lower value of AD in frame no. 50 and frame no. 450 showing better accuracy quantitatively. However, from visual results it can be clearly seen that segmented object could not show proper shape (Figs. 4*a* and *b*, from left second row fifth column) and hence this segmented frame could not be used in any shape

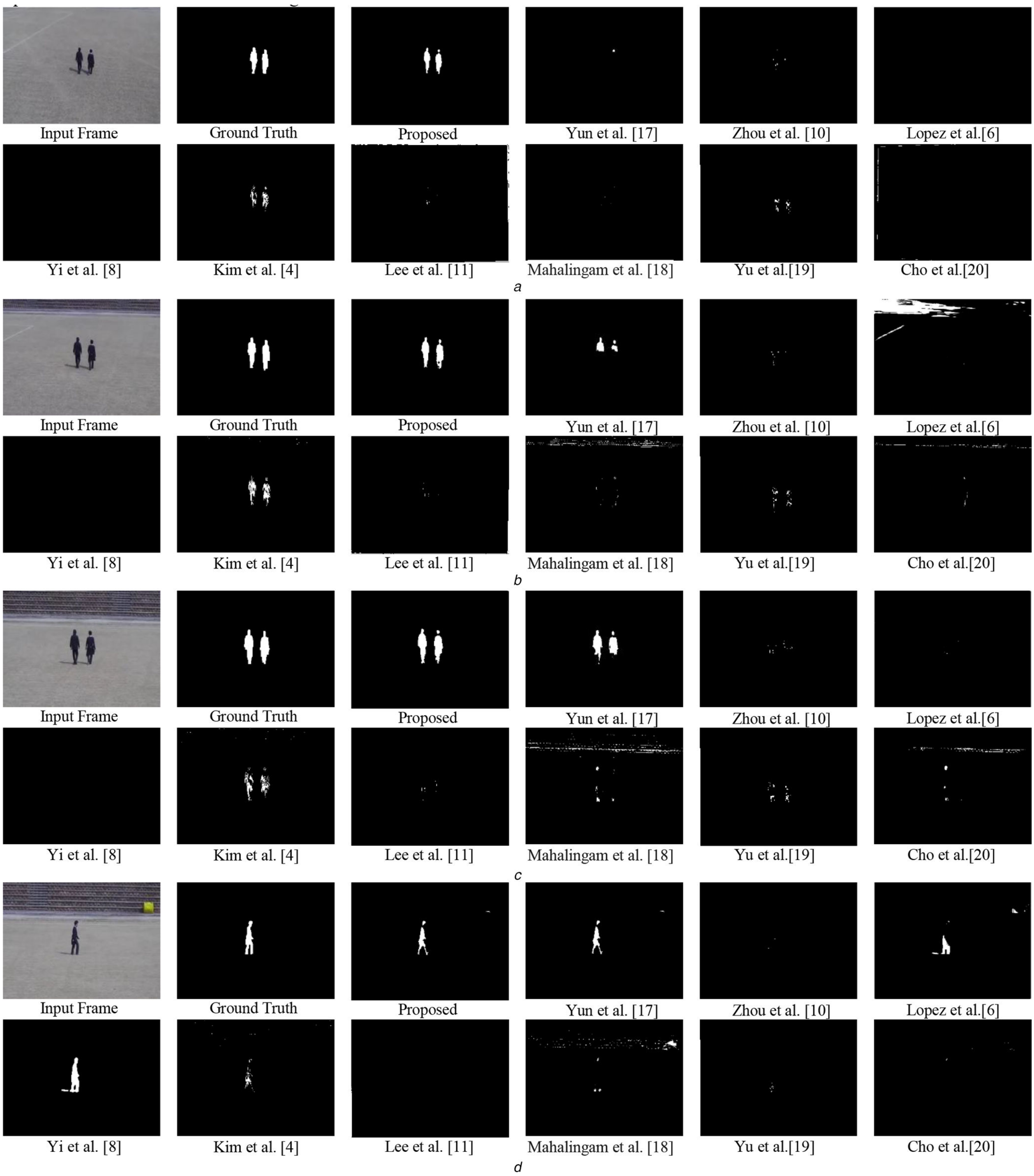


Fig. 6 *Segmentation results on a video clip of frame size 320 × 240. Frames are arranged from top to bottom as* ***(a)*** Frame no. 50, ***(b)*** Frame no. 400, ***(c)*** Frame no. 700, ***(d)*** Frame no. 900

analysis or further processing. The better segmentation performance is due to the proper consideration of motion vector of each pixel to label each pixel as a foreground or background.

5.2 Experiment #2

In this experiment, we have taken a movie clip that involves running and dancing activity of human character(s). The size of this movie frame is 640 × 272, and it contains 1158 frames with frame rate 24 fps. The proposed method and methods used in comparison [4, 6, 8, 10, 11, 17–20] have been evaluated on this video clips. The main challenge of this video clip was that object was occluded in few frames and the objects in the video were having complex activity during dancing. For representative purpose, results for frame nos. 120, 450, 950 and 1060 have been shown in Fig. 5.

From Fig. 5, one can observe that on frame no. 120 (Fig. 5*a*), object is occluded by background. Among methods including the proposed one, are not producing the results as good as ground truth. However, the proposed method is producing comparatively better segmentation of object (Fig. 5*a*, from left first row third column) than the other state-of-arts methods used in comparison [4, 6, 8, 10, 11, 17–20]. The methods proposed by Zhou *et al.* [10], Lee *et al.* [11] and Yu *et al.* [19] were not able to segment even a portion of the object (Fig. 5*a*, from left first row fifth column, Fig. 5*a*, from left second row third column and Fig. 5*a*, from left

Table 1 Performance measures for the proposed and other methods for representative segmented frames

Frame no.	Method	PCC(↑)	JC(↑)	FM(↑)	QP(↑)	AD(↓)	MSE(↓)
50	Yun *et al.* [17]	0.046	0.024	0.023	2.380	0.135	0.844
	Zhou *et al.* [10]	0.037	0.019	0.018	1.910	0.167	0.858
	Lopez *et al.* [6]	0.237	0.134	0.118	13.45	01.69	0.975
	Yi *et al.* [8]	0.141	0.076	0.070	7.590	0.328	0.945
	Kim *et al.* [4]	0.379	0.234	0.189	23.430	0.423	0.978
	Lee *et al.* [11]	0.048	0.024	0.024	2.460	0.120	0.278
	Mahalingam *et al.* [18]	0.312	0.185	0.156	18.520	0.064	0.253
	Yu *et al.* [19]	0.045	0.023	0.022	2.340	**0.007**	0.379
	Cho *et al.* [20]	0.080	0.042	0.040	4.200	0.134	0.213
	proposed	**0.430**	**0.274**	**0.215**	**27.430**	0.106	**0.115**
450	Yun *et al.* [17]	0.327	0.196	0.163	19.600	0.586	0.945
	Zhou *et al.* [10]	0.011	0.005	0.005	0.593	0.371	0.539
	Lopez *et al.* [6]	0.065	0.033	0.032	3.350	02.38	0.855
	Yi *et al.* [8]	0.217	0.122	0.108	12.210	0.975	0.805
	Kim *et al.* [4]	0.504	0.337	0.252	33.730	0.645	0.970
	Lee *et al.* [11]	0.000	0.000	0.000	0.000	0.003	0.169
	Mahalingam *et al.* [18]	0.333	0.200	0.166	20.020	0.268	0.300
	Yu *et al.* [19]	0.005	0.002	0.028	0.276	**0.008**	0.179
	Cho *et al.* [20]	0.237	0.135	0.119	13.500	0.022	0.080
	proposed	**0.560**	**0.389**	**0.280**	**38.920**	0.022	**0.027**
725	Yun *et al.* [17]	0.017	0.008	0.008	0.868	0.234	0.874
	Zhou *et al.* [10]	0.002	0.001	0.001	0.100	0.049	0.628
	Lopez *et al.* [6]	0.000	0.000	0.000	0.000	0.075	0.558
	Yi *et al.* [8]	0.148	0.080	0.074	8.030	0.376	0.755
	Kim *et al.* [4]	0.422	0.267	0.211	26.760	0.588	0.940
	Lee *et al.* [11]	0.000	0.000	0.000	0.000	0.094	0.112
	Mahalingam *et al.* [18]	0.324	0.193	0.162	19.360	0.025	0.075
	Yu *et al.* [19]	0.000	0.000	0.000	0.000	0.045	0.161
	Cho *et al.* [20]	0.153	0.083	0.076	8.320	0.072	0.080
	proposed	**0.480**	**0.316**	**0.240**	**31.640**	**0.012**	**0.061**
925	Yun *et al.* [17]	0.258	0.148	0.129	14.840	0.444	0.889
	Zhou *et al.* [10]	0.003	0.002	0.002	0.159	0.095	0.345
	Lopez *et al.* [6]	0.000	0.000	0.000	0.000	02.43	0.789
	Yi *et al.* [8]	0.199	0.110	0.099	11.070	02.20	0.825
	Kim *et al.* [4]	0.560	0.389	0.280	38.940	0.904	0.950
	Lee *et al.* [11]	0.000	0.000	0.000	0.000	0.009	0.044
	Mahalingam *et al.* [18]	0.191	0.106	0.095	10.590	0.284	0.293
	Yu *et al.* [19]	0.000	0.000	0.000	0.015	0.055	0.109
	Cho *et al.* [20]	0.049	0.025	0.024	2.510	0.004	0.044
	proposed	**0.589**	**0.417**	**0.294**	**41.750**	**0.001**	**0.007**

second row fifth column), while methods proposed by Yun *et al.* [17], Lopez *et al.* [6], Yi *et al.* [8], Kim *et al.* [4], Mahalingam *et al.* [18] and Cho *et al.* [20] produced very poor segmentation (Fig. 5*a*, from left first row fourth column, Fig. 5*a*, from left first row sixth column, Fig. 5*a*, from left second row first column, Fig. 5*a*, from left second row second column, Fig. 5*a*, from left second row fourth column and Fig. 5*a*, from left second row sixth column). On the frame no.450, the methods proposed by Lopez *et al.* [6] and Yi *et al.* [8] could not perform any segmentation and other methods by Yun *et al.* [17], Zhou *et al.* [10], Kim *et al.* [4], Lee *et al.* [11], Mahalingam *et al.* [18], Yu *et al.* [19] and Cho *et al.* [20] produced the improper segmentation of object. The results on frame no. 950 indicate that all methods performed segmentation up to some extent (Fig. 5*c*), while the proposed method performed better among these. From the results on frame no. 1060 (Fig. 5*d*), we can observe visually that the methods proposed by Yi *et al.* [8] is performing quite well for segmenting one object while could not segment properly the other object (Fig. 5*d*, from left second row first column) on the same frame. This method is also suffering from false labelling problem that leads to improper segmentation. Further, from Fig. 5, one can observe that the proposed methods outperformed in almost every representative frame amongst the methods used in comparison [4, 6, 8, 10, 11, 17–20].

The quantitative analysis of the results obtained was also performed. For this, six metric (PCC, JC, FM, QP, AD, MSE) values were computed. The resulted metric values have been shown in Table 2. The best values are shown in bold for each method. From Table 2, we observe that the proposed method provided higher segmentation accuracy than the other state-of-art methods used in comparison. Thus, from visual and quantitative analysis of segmentation results, of the video clip used in Experiment 2, we found better segmentation by the use of the proposed method.

5.3 Experiment #3

A video clip of frame size 320 × 240 containing 930 frames was used. This clip presents walking of two people. The clip is part of ground3 dataset which was earlier used by Yun *et al.* [17]. The main challenge of the video is that the object present in the video are of small size. We experimented with this clip and obtained the visual and quantitative results. The segmented visuals for representative frames 50, 400, 700 and 900 of the clip are shown in Fig. 6. We have evaluated the proposed method qualitatively and quantitatively both and compared the results with other state-of-art methods proposed by Yun *et al.* [17], Zhou *et al.* [10], Lopez *et al.*

Table 2 Performance measures for the proposed and other methods for representative segmented frames

Frame no.	Method	PCC(↑)	JC(↑)	FM(↑)	QP(↑)	AD(↓)	MSE(↓)
120	Yun *et al.* [17]	0.007	0.003	0.003	0.380	0.021	0.118
	Zhou *et al.* [10]	0.034	0.015	0.015	1.540	0.068	0.069
	Lopez *et al.* [6]	0.106	0.056	0.053	5.600	0.539	0.606
	Yi *et al.* [8]	0.085	0.044	0.042	4.440	0.253	0.360
	Kim *et al.* [4]	0.464	0.302	0.232	30.200	0.030	0.034
	Lee *et al.* [11]	0.046	0.023	0.023	2.380	0.063	0.074
	Mahalingam *et al.* [18]	0.470	0.307	0.235	30.770	0.080	0.108
	Yu *et al.* [19]	0.036	0.018	0.018	1.840	0.061	0.076
	Cho *et al.* [20]	0.105	0.055	0.052	5.570	0.059	0.060
	proposed	**0.572**	**0.401**	**0.286**	**40.100**	**0.006**	**0.023**
450	Yun *et al.* [17]	0.027	0.013	0.013	1.360	0.036	0.149
	Zhou *et al.* [10]	0.000	0.000	0.000	0.000	0.060	0.060
	Lopez *et al.* [6]	0.089	0.047	0.045	4.700	0.562	0.610
	Yi *et al.* [8]	0.000	0.000	0.000	0.000	0.060	0.060
	Kim *et al.* [4]	0.190	0.105	0.095	10.510	0.047	0.049
	Lee *et al.* [11]	0.000	0.000	0.000	0.000	0.040	0.081
	Mahalingam *et al.* [18]	0.291	0.170	0.146	17.080	0.227	0.265
	Yu *et al.* [19]	0.002	0.002	0.001	0.119	0.053	0.063
	Cho *et al.* [20]	0.090	0.047	0.045	4.750	0.046	0.051
	proposed	**0.418**	**0.264**	**0.209**	**26.490**	**0.020**	**0.039**
950	Yun *et al.* [17]	0.119	0.063	0.059	6.370	0.114	0.307
	Zhou *et al.* [10]	0.122	0.065	0.061	6.530	0.219	0.219
	Lopez *et al.* [6]	0.240	0.136	0.120	13.680	0.501	0.711
	Yi *et al.* [8]	0.271	0.157	0.135	15.700	0.074	0.268
	Kim *et al.* [4]	0.447	0.288	0.223	28.850	0.126	0.131
	Lee *et al.* [11]	0.085	0.044	0.042	4.450	0.169	0.028
	Mahalingam *et al.* [18]	0.491	0.325	0.245	32.580	0.026	0.138
	Yu *et al.* [19]	0.240	0.137	0.120	13.690	0.191	0.205
	Cho *et al.* [20]	0.088	0.460	0.044	4.610	0.202	0.204
	proposed	**0.591**	**0.419**	**0.297**	**41.990**	**0.045**	**0.053**
1060	Yun *et al.* [17]	0.257	0.147	0.128	14.780	0.126	0.468
	Zhou *et al.* [10]	0.248	0.141	0.124	14.160	0.330	0.331
	Lopez *et al.* [6]	0.178	0.097	0.085	9.780	0.134	0.474
	Yi *et al.* [8]	0.003	0.001	0.001	0.177	0.383	0.423
	Kim *et al.* [4]	0.339	0.204	0.169	20.440	0.281	0.286
	Lee *et al.* [11]	0.208	0.116	0.104	11.610	0.347	0.354
	Mahalingam *et al.* [18]	0.452	0.292	0.226	29.230	0.059	0.242
	Yu *et al.* [19]	0.391	0.243	0.195	24.330	0.256	0.281
	Cho *et al.* [20]	0.042	0.021	0.021	2.150	0.368	0.371
	proposed	**0.585**	**0.413**	**0.292**	**41.360**	**0.089**	**0.102**

[6], Yi *et al.* [8], Kim *et al.* [4], Lee *et al.* [11], Mahalingam *et al.* [18], Yu *et al.* [19] and Cho *et al.* [20].

From Fig. 6 one can observe that in segmentation result obtained, most of the methods used in comparison could not segment the objects or segmented poorly. In frame nos. 50 and 400, there are two human object appearance. From Figs. 6*a* and *b*, we observe that none of the methods used in comparison except Kim *et al.* [4] was able to segment even a part of the object(s). Kim *et al.* [4] produced poor segmentation result. The method proposed by Yun *et al.* [17] has segmented only upper half body part of human objects in frame nos. 400 and 700. This is because of poor segmentation algorithms design for small objects and slow initialisation problem of background subtraction technique used. A robust design nature of the proposed method for small size to large size object resulted in accurate segmentation even in small objects (Fig. 6*a*, from left first row third column) and (Fig. 6*b*, from left first row third column). From Fig. 6*d*, one can observe that, for frame no. 900, the method proposed by Yun *et al*. [17] resulted in a segmentation quite well which is very close to the proposed method. In frame 900, the method proposed by Yi *et al.* [8] has resulted in poor segmentation (Fig. 6*d*). Thus, visually we saw that the proposed method resulted in segmentation close to the ground truth in almost every frame of this clip.

The quantitative assessment of the results was performed using six performance measures viz. PCC, JC, FM, QP, AD and MSE. The computed values of performance measures have been given in Table 3. The best values are shown in bold for each method. From Table 3, one can observe that, for frame nos. 700 and 900, Yi *et al.* [8] resulted in higher values of PCC, JC, FM, QP and lower values of AD and MSE, which mean better segmentation. However, from visual results we can see that it resulted in distorted segmented object. This is due to segmenting shadow as part of object. Thus, from the visual results (Fig. 6) as well as quantitative measures (Table 3), we found that the proposed method has performed better and produced results close to the ground truth in some frames.

In Figs. 4–6 and Tables 1–3, we presented segmentation results of the proposed method and its comparison with other state-of-the-art methods. The experiments were performed on several video clips with different complex real scenarios like presence of abrupt change in illumination, small object size, large object size, low and high speed of the object, ceased object and so on. From visual and quantitative analyses, we observed that the proposed method outperformed over other state-of-the-art methods. The outperformance is due to the use of dense optical flow for accurate identification of moving pixels. Dense optical flow gives the displacement of each pixel in between consecutive frames.

Table 3 Performance measures for the proposed and other methods for representative segmented frames

Frame no.	Method	PCC(↑)	JC(↑)	FM(↑)	QP(↑)	AD(↓)	MSE(↓)
50	Yun *et al.* [17]	0.030	0.015	0.015	1.530	0.012	0.013
	Zhou *et al.* [10]	0.023	0.011	0.011	1.170	0.011	0.012
	Lopez *et al.* [6]	0.000	0.000	0.000	0.000	0.547	0.573
	Yi *et al.* [8]	0.004	0.024	0.024	0.238	0.012	0.012
	Kim *et al.* [4]	0.428	0.272	0.214	27.230	0.000	0.007
	Lee *et al.* [11]	0.021	0.010	0.021	1.060	0.011	0.039
	Mahalingam *et al.* [18]	0.005	0.002	0.002	0.294	0.009	0.017
	Yu *et al.* [19]	0.132	0.070	0.066	7.080	0.032	0.020
	Cho *et al.* [20]	0.000	0.000	0.000	0.000	0.083	0.019
	proposed	**0.556**	**0.387**	**0.278**	**38.560**	**0.003**	**0.006**
400	Yun *et al.* [17]	0.387	0.240	0.192	24.010	0.009	0.011
	Zhou *et al.* [10]	0.028	0.014	0.014	1.450	0.016	0.017
	Lopez *et al.* [6]	0.000	0.000	0.000	0.000	0.543	0.577
	Yi *et al.* [8]	0.000	0.000	0.000	0.000	0.017	0.018
	Kim *et al.* [4]	0.336	0.202	0.168	20.210	0.008	0.011
	Lee *et al.* [11]	0.009	0.004	0.009	0.446	0.007	0.030
	Mahalingam *et al.* [18]	0.001	0.000	0.000	0.078	0.019	0.019
	Yu *et al.* [19]	0.077	0.040	0.038	4.010	0.010	0.024
	Cho *et al.* [20]	0.000	0.000	0.000	0.000	0.013	0.025
	proposed	**0.495**	**0.329**	**0.248**	**32.900**	**0.004**	**0.008**
700	Yun *et al.* [17]	**0.591**	**0.419**	**0.295**	**41.950**	**0.001**	**0.007**
	Zhou *et al.* [10]	0.037	0.019	0.018	1.930	0.020	0.021
	Lopez *et al.* [6]	0.000	0.000	0.000	0.000	0.539	0.581
	Yi *et al.* [8]	0.000	0.000	0.000	0.000	0.022	0.022
	Kim *et al.* [4]	0.423	0.268	0.211	26.810	0.006	0.010
	Lee *et al.* [11]	0.025	0.012	0.025	1.292	0.023	0.023
	Mahalingam *et al.* [18]	0.084	0.044	0.041	4.370	0.002	0.044
	Yu *et al.* [19]	0.094	0.049	0.047	4.930	0.013	0.027
	Cho *et al.* [20]	0.059	0.030	0.029	3.061	0.011	0.032
	proposed	0.573	0.402	0.287	40.190	**0.001**	**0.007**
900	Yun *et al.* [17]	**0.616**	**0.445**	**0.308**	**44.570**	**0.001**	**0.002**
	Zhou *et al.* [10]	0.006	0.003	0.003	0.339	0.008	0.008
	Lopez *et al.* [6]	0.005	0.002	0.003	0.250	0.556	0.569
	Yi *et al.* [8]	0.608	0.437	0.304	43.760	0.005	0.005
	Kim *et al.* [4]	0.308	0.182	0.156	18.240	0.003	0.006
	Lee *et al.* [11]	0.000	0.000	0.000	0.000	0.008	0.011
	Mahalingam *et al.* [18]	0.083	0.043	0.041	4.320	0.002	0.020
	Yu *et al.* [19]	0.013	0.007	0.006	0.696	0.002	0.016
	Cho *et al.* [20]	0.012	0.006	0.062	0.619	0.009	0.010
	proposed	0.450	0.290	0.225	29.080	**0.001**	0.006

Consideration of motion vector of each pixel makes result more promising. The proposed method is robust in the sense that it could be used for different object sizes, varying lighting, occlusion and in presence of shadow.

6 Conclusions

In this paper, we proposed a dense optical flow based method for object segmentation in video with moving camera scenario. The algorithm is useful for object segmentation in several realistic videos. In the proposed work, both camera motion and object motion have been estimated separately. Firstly, camera motion was compensated by geometric computation. Then a statistical single Gaussian approach was used to model the background and its updation. The proposed method is new in the sense that the foreground pixels were identified by integration of background subtraction and dense optical flow. For correct identification of foreground and background pixels, both foreground and background pixels were labelled and then foreground pixels were again investigated for the deciding whether pixel falls in foreground or background using magnitude of motion flow vector. The segmentation results of the proposed work were evaluated and compared qualitatively and quantitatively with the methods: Yun *et al.* [17], Zhou *et al.* [10], Lopez *et al.* [6], Yi *et al.* [8], Kim *et al.* [4], Lee *et al.* [11], Mahalingam *et al.* [18], Yu *et al.* [19] and Cho *et al.* [20]. Quantitative evaluation was performed using six quantitative metrics – PCC, JC, FM, QP, AD and MSE. The proposed method performs better and has better object shape preserving capability than the other state-of-art methods used in comparison. From the series of experiments and its evaluations, we observed that the proposed method outperformed for segmenting the object in video with moving camera. Also, we observed that the proposed method is capable of performing the segmentation in realistic unconstrained video having object size from small to large. The proposed method works well in cluttered and varying lighting conditions.

7 References

[1] Zhang, H.B, Zhang, Y.X, Zhong, B, *et al.*: 'A comprehensive survey of vision-based human action recognition methods', *Sensors*, 2019, **19**, (5), p. 1005

[2] Ke, S.R, Thuc, H., Lee, Y.J., *et al.*: 'A review on video-based human activity recognition', *Computers*, 2013, **2**, (2), pp. 88–131

[3] Yazdi, M., Bouwmans, T.: 'New trends on moving object detection in video images captured by a moving camera: A survey', *Comput. Sci. Rev.*, 2018, **28**, pp. 157–177

[4] Kim, S.W., Yun, K., Yi, M.K., *et al.*: 'Detection of moving objects with a moving camera using non-panoramic background model', *Mach. Vis. Appl.*, 2013, **24**, (5), pp. 1015–1028

[5] Hu, W.C., Chen, C.H., Chen, T.Y., *et al.*: 'Moving object detection and tracking from video captured by moving camera', *J. Vis. Commun. Image Represent.*, 2015, **30**, pp. 164–180

[6] López-Rubio, F.J., López-Rubio, E.: 'Foreground detection for moving cameras with stochastic approximation', *Pattern Recognit. Lett.*, 2015, **68**, pp. 161–168

[7] Kim, J., Wang, X., Wang, H., *et al.*: 'Fast moving object detection with non-stationary background', *Multimedia Tools Appl.*, 2013, **67**, (1), pp. 311–335

[8] Yi, K.M., Yun, K., Wan Kim, S.J., *et al.*: 'Detection of moving objects with non-stationary cameras in 5.8 ms: bringing motion detection to your mobile device'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 2013, pp. 27–34

[9] Kushwaha, A., Prakash, O., Srivastava, R.K., *et al.*: 'Dense flow-based video object segmentation in dynamic scenario'. Recent Trends in Communication, Computing, and Electronics, Singapore, 2019, pp. 271–278

[10] Zhou, D., Frémont, V., Quost, B., *et al.*: 'Moving object detection and segmentation in urban environments from a moving platform', *Image Vision Comp.*, 2017, **68**, pp. 76–87

[11] Lee, S., Kim, N., Jeong, K., *et al.*: 'Moving object detection using unstable camera for video surveillance systems', *Optik*, 2015, **126**, (20), pp. 2436–2441

[12] Fleet, D., Weiss, Y.: 'Optical flow estimation'. Handbook of mathematical models in computer vision, Boston MA, 2006, pp. 237–257

[13] Horn, B.K., Schunck, B.G.: 'Determining optical flow', *Artif. Intell.*, 1981, **17**, (1–3), pp. 185–203

[14] Baghaie, A., D'Souza, R., Yu, Z.: 'Dense descriptors for optical flow estimation: a comparative study', *J. Imaging*, 2017, **3**, (1), pp. 1–19

[15] Anandan, P.: 'A computational framework and an algorithm for the measurement of visual motion', *Int. J. Comput. Vis.*, 1989, **2**, (3), pp. 283–310

[16] Kurnianggoro, L., Shahbaz, A., Jo, K.H.: 'Dense optical flow in stabilized scenes for moving object detection from a moving camera'. 16th Int. Conf. on Control, Automation and Systems (ICCAS), Gyeongju, Republic of Korea, October 2016, pp. 704–708

[17] Yun, K., Lim, J., Choi, J.Y.: 'Scene conditional background update for moving object detection in a moving camera', *Pattern Recognit. Lett.*, 2017, **88**, pp. 57–63

[18] Mahalingam, T., Subramoniam, M.: 'A robust single and multiple moving object detection, tracking and classification', *Appl. Comput. Inf.*, 2018, to appear, doi: 10.1016/j.aci.2018.01.001

[19] Yu, Y., Kurnianggoro, L., Jo, K.H.: 'Moving object detection for a moving camera based on global motion compensation and adaptive background model', *Int. J. Control, Autom. Syst.*, 2019, **17**, (7), pp. 1866–1874

[20] Cho, J., Jung, Y., Kim, D.S., *et al.*: 'Moving object detection based on optical flow estimation and a Gaussian mixture model for advanced driver assistance systems', *Sensors*, 2019, **19**, (14), p. 3217

[21] Mishra, A.K., Aloimonos, Y., Cheong, L.F., *et al.*: 'Active visual segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (4), pp. 639–653

[22] Andriluka, M., Roth, S., Schiele, B.: 'People-tracking-by-detection and people-detection-by-trackin'. IEEE Conf. on computer vision and pattern recognition, Anchorage, AK, USA, June 2008, pp. 1–8

[23] Li, X., Xu, C.: 'Moving object detection in dynamic scenes based on optical flow and superpixels'. IEEE Int. Conf. on Robotics and Bio. (ROBIO), Zhuhai, People's Republic of China, 2015, pp. 84–89

[24] Huang, J., Zou, W., Zhu, J., *et al.*: 'Optical flow based real-time moving object detection in unconstrained scenes', arXiv preprint arXiv, 1807.04890, July 2018

[25] Stauffer, C., Grimson, W.E.: 'Adaptive background mixture models for real-time tracking'. IEEE Computer Society Conf. on Computer Vision and Pattern Recog., Fort Collins, CO, USA, June 1999, vol. 2, pp. 246–252

[26] Wren, C.R., Azarbayejani, A., Darrell, T., *et al.*: 'Pfinder: real-time tracking of the human body', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 780–785

[27] Bouwmans, T., El Baf, F., Vachon, B.: 'Background modeling using mixture of Gaussians for foreground detection-a survey', *Recent Patents Comput. Sci.*, 2008, **1**, (3), pp. 219–237

[28] Khare, M, Srivastava, R.K., Khare, A.: 'Moving object segmentation in daubechies complex wavelet domain', *Signal, Image Video Process.*, 2015, **9**, (3), pp. 635–650

[29] Khare, M., Srivastava, R.K., Khare, A., *et al.*: 'Single change detection-based moving object segmentation by using daubechies complex wavelet transform', *IET Image Proc.*, 2014, **8**, (6), pp. 334–344

[30] Mittal, A., Huttenlocher, D.: 'Scene modeling for wide area surveillance and image synthesis'. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head Island, SC, USA, June 2000, vol. 2, pp. 160–167

[31] Cucchiara, R., Prati, A., Vezzani, R., *et al.*: 'Advanced video surveillance with pan tilt zoom cameras'. Proc. of the 6th IEEE Int. Workshop on Visual Surveillance, Graz, Austria, May 2006, pp. 334–352

[32] Cho, S.H., Kang, H.B.: 'Panoramic background generation using mean-shift in moving camera environment'. Proc. of the international conference on image processing, computer vision, and pattern recognition (IPCV), 2011, pp. 1–7

[33] Xue, K., Liu, Y., Ogunmakin, G., *et al.*: 'Panoramic Gaussian mixture model and large-scale range background substraction method for PTZ camera-based surveillance systems', *Mach. Vis. Appl.*, 2013, **24**, (3), pp. 477–492

[34] Lucas, B.D., Kanade, T.: 'An iterative image registration technique with an application to stereo vision'. Proc. DARPA Image Understanding Workshop, Vancouver, BC, Canada, April 1981, pp. 121–130

[35] Azzari, P., Bevilacqua, A.: 'Joint spatial and tonal mosaic alignment for motion detection with ptz camera'. In Int. Conf. Image Analysis and Recognition, Berlin, Heidelberg, September 2006, pp. 764–775

[36] Lowe, D.G.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110

[37] Dubrofsky, E.: 'Homography estimation', Diplomovápráce Vancouver, UniverzitaBritskéKolumbie, March 2009

[38] Fischler, M.A., Bolles, R.C.: 'Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography', *Commun. ACM*, 1981, **24**, (6), pp. 381–395

[39] Adelson, E.H., Bergen, J.R.: 'Spatiotemporal energy models for the perception of motion', *Josa a.*, 1985, **2**, (2), pp. 284–299

[40] Farnebäck, G.: 'Two-frame motion estimation based on polynomial expansion'. Scandinavian Conf. on Image Analysis, Berlin, Heidelberg, 2003, pp. 363–370

[41] Harris, C.G.: 'Stephens M. A combined corner and edge detector', *InAlvey Vis. Conf.*, 1988, **15**, (50), pp. 147–151

[42] Tuytelaars, T., Mikolajczyk, K.: 'Local invariant feature detectors: a survey', *Found. Trends® Comput. Graph. Vis.*, 2008, **3**, (3), pp. 177–280

[43] Bommisetty, R.M., Prakash, O., Khare, A.: 'Video superpixels generation through integration of curvelet transform and simple linear iterative clustering', *Multimed. Tools Appl.*, 2019, **78**, (17), pp. 25185–25219